

IN THE CLAIMS:

Please amend the claims as indicated below:

1. (Currently Amended) A method comprising the steps of:

providing a set of sequences, wherein:

the sequences are not aligned; and

5 each sequence comprises a series of symbols;

discovering a plurality of patterns common to a plurality of the sequences,

wherein each pattern comprises a plurality of positions, at least one of the positions comprise an expected symbol and at least one of the positions comprise one symbol of a specified plurality of symbols, wherein the specified plurality of symbols consists of at least two

10 symbols and no more than $|\Sigma|-1$ symbols, wherein $|\Sigma|$ is a number of available symbols in a set of amino acids, and wherein Σ consists of A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, I, V, W and Y; and

determining if a candidate sequence comprises a predetermined number of the patterns.

15

2. (Original) The method of claim 1, wherein the patterns common to a plurality of the set of sequences comprise test patterns, wherein the sequences in set of sequences comprise test sequences, and wherein the step of determining if a candidate sequence comprises a predetermined number of the patterns comprises the step of determining if there are

20 candidate patterns in the candidate sequence that match all of the predetermined number of test patterns.

3. (Original) The method of claim 1, further comprising the step of determining if each of the plurality of patterns is statistically significant.

25

4. (Previously Presented) The method of claim 1, wherein the step of discovering is performed without using any knowledge about biological information related to family, cardinality or image characteristics of sequences in the set of unaligned sequences.

5 (Previously Presented) The method of claim 1, further comprising the steps of:

if the candidate sequence comprises the predetermined number of patterns,
adding the candidate sequence to the set of sequences to create a new set of sequences; and

5 performing the step of discovering on the new set of sequences

6 (Previously Presented) The method of claim 1, wherein some of the plurality of
positions comprise positions which may be occupied by any sequence character

10 7. (Cancelled)

8 (Original) The method of claim 3, wherein the step of determining if each of the
plurality of patterns is statistically significant comprises the steps of selecting one of the
patterns, determining if a probability that the selected pattern occurs in a sequence meets a
15 predetermined threshold, and continuing to select additional patterns until each pattern has
been selected.

9 (Original) The method of claim 8, wherein the step of determining if a probability
that the selected pattern occurs in a sequence meets a predetermined threshold further
20 comprises the steps of using a second-order Markov chain method to determine the
probability that the selected pattern occurs in a sequence and determining a natural logarithm
of the probability that the selected pattern occurs in a sequence.

10 (Original) The method of claim 3, wherein the step of determining if each of the
25 plurality of patterns is statistically significant further comprises the steps of removing
instances of each of the patterns from the set of sequences to create a new set of sequences
and performing the step of discovering on the new set of sequences.

11 (Original) The method of claim 3, wherein the step of determining if each of the
30 plurality of patterns is statistically significant further comprises the steps of if any of the
patterns is statistically significant, selecting a statistically significant pattern, modifying a

composite descriptor to include the selected pattern if the selected pattern is not already part of the composite descriptor, and continuing to select statistically significant patterns until all statistically significant patterns have been selected.

5 12 (Original) The method of claim 1, wherein the step of discovering a plurality of patterns common to a plurality of the sequences comprises the steps of:

selecting a predetermined threshold that indicates how many of the sequences should contain a pattern for the pattern to be considered common;

10 of sequences;
 discovering patterns, if any, that are common to the predetermined threshold

if there are no patterns common to the predetermined threshold of sequences, decreasing the predetermined threshold; and

performing, until the predetermined threshold is less than a predetermined amount, the step of discovering patterns, if any, that are common to the predetermined
 15 threshold of sequences and the step of if there are no patterns common to the predetermined threshold of sequences, decreasing the predetermined threshold.

13. (Canceled)

20 14. (Canceled)

15. (Canceled)

25 16. (Canceled)

17. (Canceled)

18. (Canceled)

30 19. (Canceled)

20 (Canceled)

21 (Canceled)

5 22 (Canceled)

23 (Currently Amended) A system comprising:

a memory that stores computer-readable code; and

a processor operatively coupled to said memory, said processor configured to

10 implement said computer-readable code, said computer-readable code configured to:

provide a set of sequences, wherein:

the sequences are not aligned; and

each sequence comprises a series of symbols;

15 discover a plurality of patterns common to a plurality of the
sequences, wherein each pattern comprises a plurality of positions, at least one of the
positions comprise an expected symbol and at least one of the positions comprise one
symbol of a specified plurality of symbols, wherein the specified plurality of symbols
consists of at least two symbols and no more than $|\Sigma| - 1$ symbols, wherein $|\Sigma|$ is a number of
available symbols in a set of amino acids, and wherein Σ consists of A, C, D, E, F, G, H, I,
20 K, L, M, N, P, Q, R, S, T, V, W and Y; and

determine if a candidate sequence comprises a predetermined number
of the patterns.

24 (Canceled)

25

25 (Currently Amended) An article of manufacture comprising:

a computer readable medium having computer readable code means embodied
thereon, said computer readable program code means comprising:

a step to provide a set of sequences, wherein:

30 the sequences are not aligned; and

each sequence comprises a series of symbols;

a step to discover a plurality of patterns common to a plurality of the sequences, wherein each pattern comprises a plurality of positions, at least one of the positions comprise an expected symbol and at least one of the positions comprise one
5 symbol of a specified plurality of symbols, wherein the specified plurality of symbols consists of at least two symbols and no more than $|\Sigma|-1$ symbols, wherein $|\Sigma|$ is a number of available symbols in a set of amino acids, and wherein Σ consists of A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W and Y; and

a step to determine if a candidate sequence comprises a predetermined
10 number of the patterns.

26. (Canceled)